REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	Oct 94	Final 1 Jul 91-30 Sep 94

4. TITLE AND SUBTITLE

Topics in Exploratory and Speculative Data Analysis

5. FUNDING NUMBERS

DAAL03-91-G-0210

6. AUTHOR(S)

James R. Thompson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Rice University
Department of Statistics
Houston, TX 77251-1892

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

ARO 28686.3-MA

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Synopses are presented for nine published papers generated by this research. Personnel supported by this research and doctoral degrees are listed.

DTIC ELECTE FEB 0 6 1994 F

DTIC QUALITY INSPECTED 4

19950202 096

14. SUBJECT TERMS	15. NUMBER OF PAGES
Data Analysis, Tumor Metastases, Stochastic Processes, Structural Failure, Design, Data Sets, Personal Computers, Multivariate Control, Quality Control, Sampling	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Final Report

Topics in Exploratory and Speculative Data Analysis

CONTRACT OR GRANT NUMBER: DAAL-03-91-G-0210

AUTHOR OF REPORT: James R. Thompson

Department of Statistics, Rice University, Houston, Texas
77251-1892

Outline of Results

The Brown-Milas-Thompson paper consists of analysis of experimental data in the light of earlier Brown-Thompson work on the influence of tumor age in its propensity to throw off metastases. The notion of tumor metastatic propensity due to tumor size is also investigated.

The Capps-Thompson paper examines the relationship between stochastic models of structural failure and the empirical design safety factors currently in use. The value of a safety factor is related to the probability of structural failure.

Marc Elliott worked with Thompson on algorithms for finding local modes of higher dimensional data sets. Various contaminated distributions were examined in attempts to stress the mean update algorithm employed. This work is leading to a doctoral dissertation which is scheduled to be completed by Elliott in 1995.

The Ensor-Bridges-Lawera paper addresses the classical homogeneous birth and death process. The computer intensive tasks have been speeded up by parallelization on a desktop LEVCO parallel system.

In the Ensor-Bridges-Thompson paper, we have considered the personal computer market, in which PC clones regularly appear and disappear, according to a nonstandard birth and death mechanism. The axioms of the model are straightforward, but the resulting likelihood function defies tractability. The use of SIMEST enabled satisfactory estimation of the characterizing parameters using as the data base the actual market entrances and exits in the PC market since its inception. Parallel computation is utilized to facilitate the intensive computing tasks using a LEVCO transputer system.

In the first of the two Lawera-Thompson papers, new procedures for multivariate control charting were developed, with emphasis on robustness. Lawera's "King of the Mountain" algorithm shows particular promise as a procedure for estimating the dominant mean with the kinds of contamination found in multivariate quality con-

Accesion For
NTIS CRA&I
DTIC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

trol data. Two rank tests were developed for use in finding out of control epochs in situations where the data is multivariate and may not be Gaussian.

In the second Lawera-Thompson paper, the SIMEST algorithm for obtaining estimates of the parameters characterizing a stochastic process is implemented using a piecewise quadratic approximation to a goodness of fit statistic. The implementation is motivated in part by the rotatable experimental designs of Box and Hunter. Here, however, an "experiment" is simply a computer simulation, so the cost of the experiment is, essentially, trivial. Parallelized computation is used on a Levco transputer system utilizing design points in a fashion so as to maximize the utilization of all transputers.

The Taylor-Thompson paper gives a new look at the nonparametric density estimation based resampling algorithm of the authors, first published in 1982. Some advantages over the bootstrap algorithm are listed. It is argued that the SIMDAT algorithm is the natural way for resampling in most situations.

Statistical Process Control for Quality Improvement by Thompson and Koronacki is a model based analysis of statistical process control. The work demonstrates the importance of the philosophy of Vilfredo Pareto to an understanding of the reasons why the Deming paradigm,, frequently dubbed "statistical process control" works. Output optimization is generally a linear control problem. System optimization is generally a nonlinear control problem. SPC is simply a step-wise means of solving the nonlinear control problem. Case study examples, based in large measure on years of field experience of the authors, are considered. Particular emphasis is given to considerations of multivariate data streams and robust procedures for their analysis.

Personnel Supported

Elliott, Marc: Graduate Student, Department of Statistics
Ensor, Katherine B.: Associate Professor of Statistics
Koronacki, Jacek: Visiting Associate Professor of Statistics

Lawera, Martin: Graduate Student, Department of Statistics
Spears, Marty: Graduate Student, Department of Statistics
Thompson, James R.: Professor of Statistics
West, Ronnie W.: Graduate Student, Department of Statistics

Doctoral Degrees Awarded to Supported Students

Spears, Floyd W. (1992) *Multi-Stage Designs in Dose-Response Studies.*

Ronnie W. West (1994), *Modeling the Potential Impact of HIV on the Spread of Tuberculosis in the United States.*

Publications

Brown, Barry W., Milas, Luka and Thompson, James R. (1992) "Effects of size and growth time of a murine sarcoma on its metastatic spread." (1992) *Clinical and Experimental Metastasis*, v. 10, pp. 77-86.

Capps, R.W. and Thompson, James R. (1993), "Statistical safety factors reduce overdesign," in *Hydrocarbon Processing*, pp. 77-82.

Elliott, Marc and Thompson, James R. (1993) "An exploratory algorithm for the estimation of mode location and numerosity in multidimensional data," in *Proceedings of the Thirty-Eighth Conference on the Design of Experiments in Army Research Development and Testing*, pp. 229-244.

Ensor, Katherine, Bridges, Eileen and Lawera, Martin. "Simulation based estimation for birth and death processes," *Proceedings of the Thirty-Seventh Conference on the Design of Experiments in Army Research Development and Testing.*

Ensor, Katherine B., Bridges, Eileen, and Thompson, James R. (1992) "Marketplace competition in the personal computer industry," *Decision Sciences*, pp. 467-477.

Lawera, Martin and Thompson, James R. (1993) "Some problems of estimation and testing in multivariate statistical process control," in *Proceedings of the Thirty-Eighth Conference on the Design of Experiments in Army Research Development and Testing*, pp. 99-126.

Lawera, Martin and Thompson, James R. (1993) "A parallelized, simulation based algorithm for parameter estimation," in *Proceedings of the Thirty-Eighth Conference on the Design of Experiments in Army Research Development and Testing*, pp. 321-341.

Taylor, Malcolm S. and Thompson, James R. (1992) "A nonparametric density estimation based resampling algorithm." (1992), in *Exploring the Limits of the Bootstrap*, R. LePage and L. Billard, eds., New York: John Wiley & Sons, pp. 397-403.

Thompson, James R. and Koronacki, Jacek (1993). *Statistical Process Control for Quality Improvement*, New York: Chapman & Hall, 391 pages.